**Comparative Analysis of Machine Learning Models for GDP Forecasting: A Case Study on the Indian Economy**

Mohana Krishnan, Suresh Nepal

Department of Mathematics, School of Advance Sciences

Vellore Institute of Technology, Vellore – 632014, Tamil Nadu, India.

E-mail: [suresh.nepal2022@vitstudent.ac.in](mailto:suresh.nepal2022@vitstudent.ac.in), [mohanakrishnan.v2022@vitstudent.ac.in](mailto:mohanakrishnan.v2022@vitstudent.ac.in)

***Abstract—*** ***In this study, the use of machine learning algorithms to anticipate GDP per capita in India is examined. We applied the XGBoost regressor, Support Vector Regressor, Random Forest Regressor, Long Short-Term Memory (LSTM) deep learning model, and ARIMA forecasting model in order to allow thorough analysis. Leveraging the Root Mean Square Error (RMSE) number, we evaluate these models' accuracy***. ***The Random Forest Regressor and ARIMA forecasting models performed more effectively than the other models and have the best accuracy when it comes to predicting India's GDP per capita. This study enhances economic predictive modeling and offers insightful information to educators, economists, and policymakers concerning GDP forecasting and India's economic development.***

*Keywords: GDP per capita, Forecasting, Machine learning models, ARIMA, Root Mean Square Error (RMSE)*

1. INTRODUCTION

Gross Domestic Product (GDP) per capita stands as a pivotal indicator in the realm of economics, serving as a yardstick to gauge a nation's economic well-being and the standard of living of its citizens. As the fundamental measure of a country's economic performance, forecasting GDP per capita plays a crucial role in policymaking, investment decisions, and socioeconomic planning.

In the pursuit of accurate GDP per capita forecasting, we assess several machine learning models.

The present research was inspired by the ground breaking work of Himani Patil, Soniya Gawade, and Tanvi Gharte, outlined in their release "GDP Prediction and Forecasting using Machine Learning." This study has solid foundation thanks to their thorough review of social, economic, and cultural factors covering the years 1970–2018 and the adoption of techniques for supervised learning. In their investigation [1], the writers contrasted the functionality of their models using three distinct techniques. Ultimately concluding that Gradient Boosting, Random Forest and Linear Regression were the best for the job.

This research strives to offer a more comprehensive and robust approach to GDP per capita forecasting rather that GPD forecasting as carried out by Himani Patil and Soniya Gawade, contributing to a deeper understanding of economic dynamics and allowing more informed and data-driven decision-making. It constructs upon the work of Gharte, Patil, and Gawade and determines a wider array of machine-learning models.

1. METHODOLOGY

***Data collection:*** The World Bank website offered data on India's GDP per capita from 1990 to 2022. The data provided covered the growth rate and the actual GDP per capita.

***Data preparation***: We conducted an initial exploratory analysis in order to look for any outliers or missing values. Therefore, there were no abnormalities in the data.

***Data splitting:*** The data was split into a training set and a testing set, with a split ratio of 80:20.

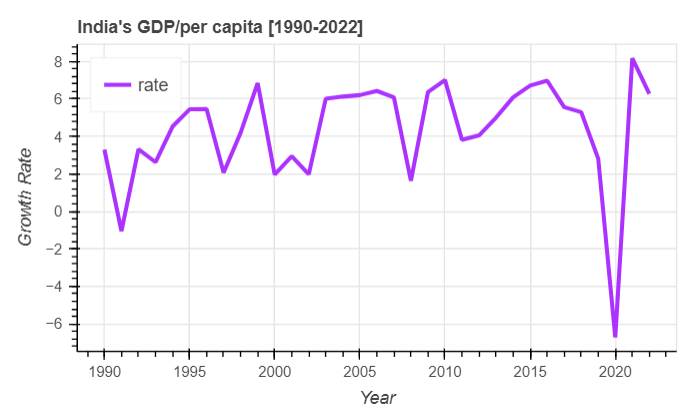
***Model selection:*** Five different models were selected for the study – ARIMA, SVMR, LSTM, Random Forest Regressor and XGBoost Regressor.

***Model implementation:*** The selected models were implemented using Python programming language and Jupyter Notebook.

***Model evaluation:*** The trained models have been tested using the testing set. Every model has been used for making predictions, and the accuracy of each model was examined by employing the Root Mean Squared Error.

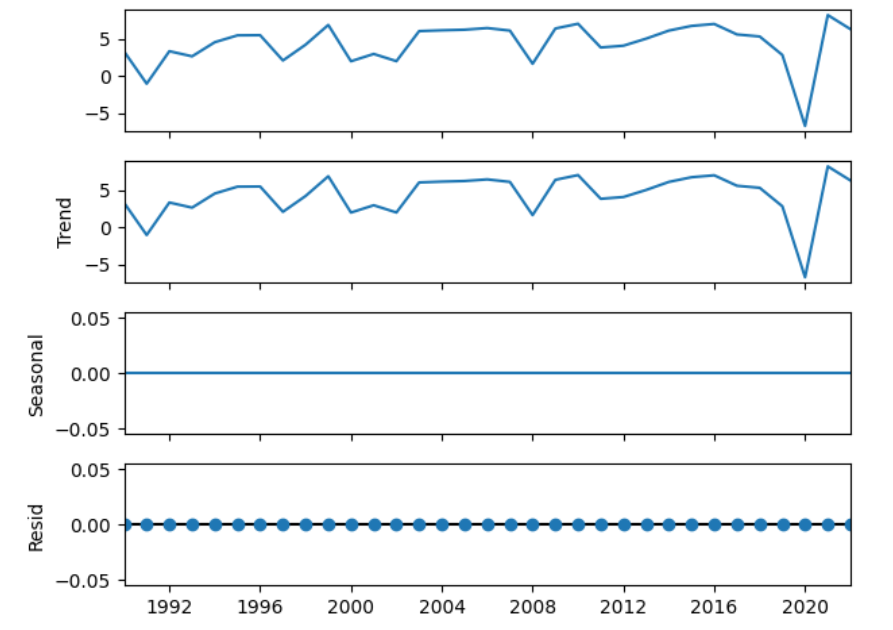
***Results analysis:*** The examinations of the models' outcomes went on so as to discover which model had the greatest accuracy score.

The actual time series data plot is presented inFigure 1.



**Figure 1:** Plot of the time series data

Figure 2shows the decomposition of the time series data into trend, seasonal and residual component after differencing

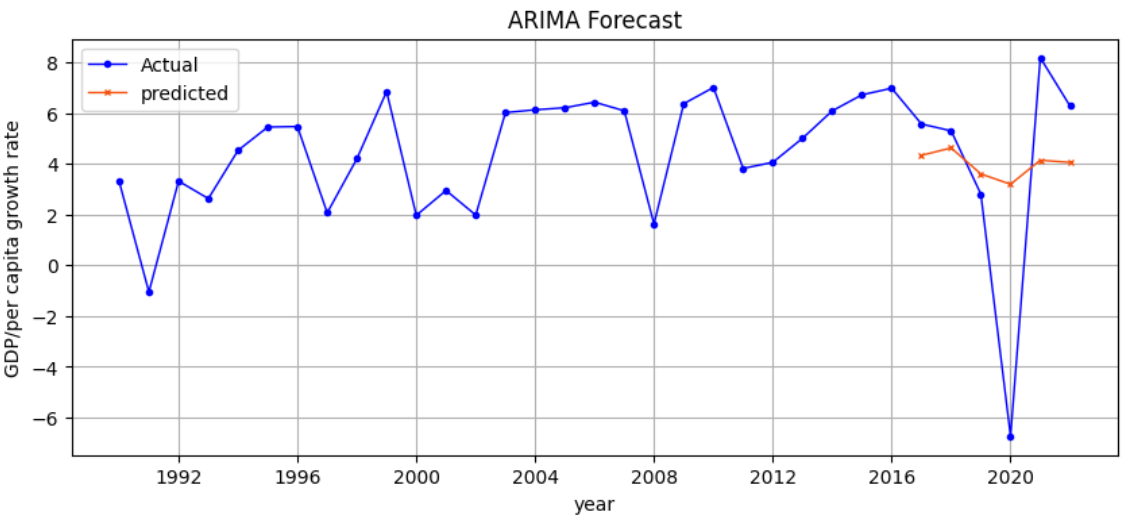


**Figure 2**: Decomposition of the time series data

# ARIMA

Based on past information, time series forecasting approaches such as ARIMA (Autoregressive Integrated Moving Average) are used for estimating future values. The data's stationarity was confirmed, and it is known ARIMA model only fits stationary data. For the ARIMA model, the Root Mean Squared Error (RMSE) was 4.51.

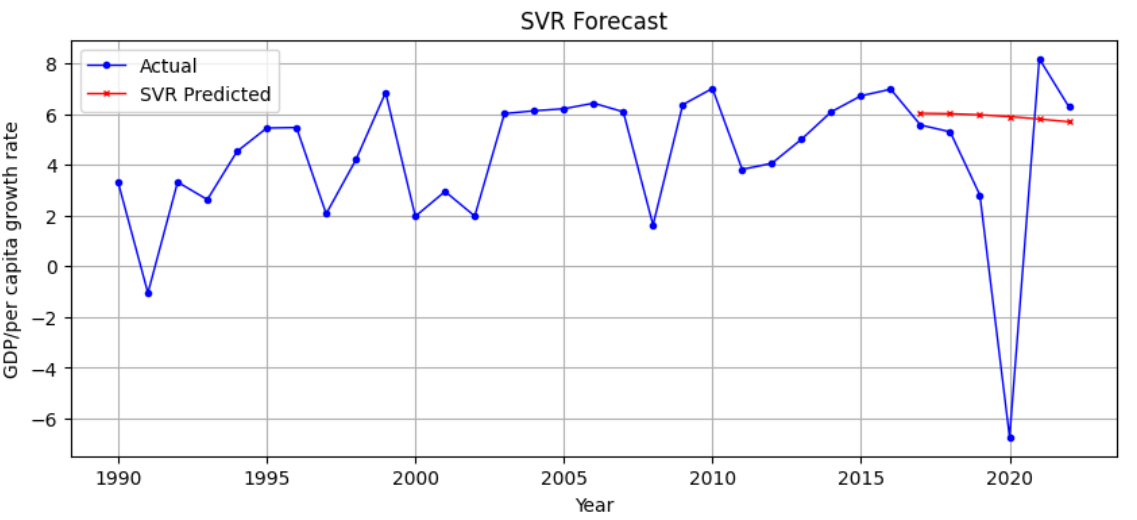
Figure 3 illustrates the actual and ARIMA-predicted data.



**Figure 3:** Actual data and predicted data using ARIMA model

# IV SUPPORT VECTOR REGRESSOR

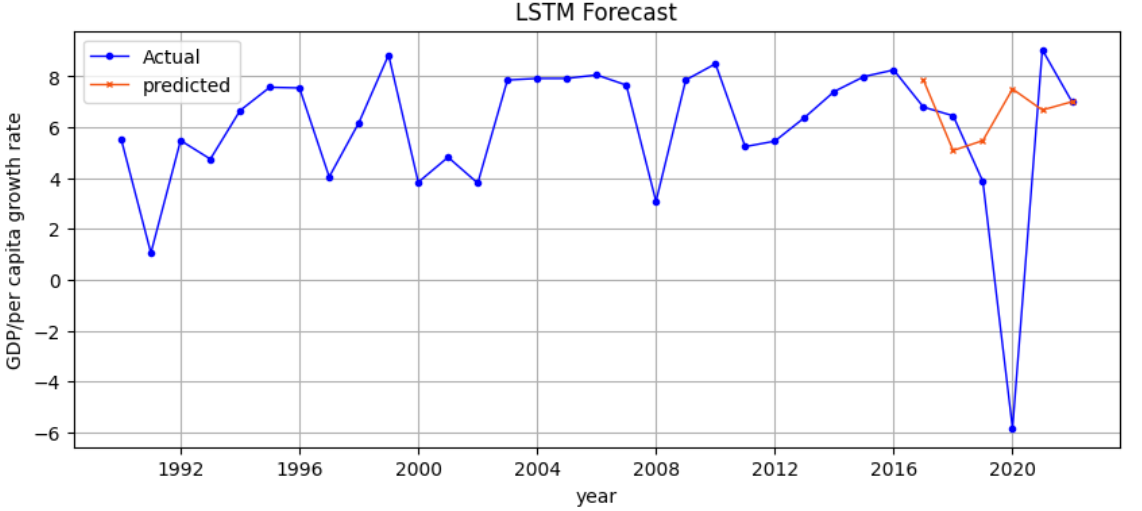
One machine-learning approach for regression tasks is Support Vector Regressor (SVR). To determine the optimal representation of the connection between the input features and the goal variable, a hyperplane in a high-dimensional feature space is identified. SVR produced an RMSE score of 5.42. Actual and SVR model-predicted data appears in Figure 3.



**Figure 4:** Actual data and predicted data using SVR model

# V LONG SHORT TERM MEMORY

Long Short Term Memory (LSTM), a type of Recurrent neural networks (RNNs), are popular for having the ability to model complicated sequences and long-term relationships in time series data. In time series forecasting, LSTMs are used to predict future values from previous information. The LSTM model's estimated RMSE rating was 5.61.

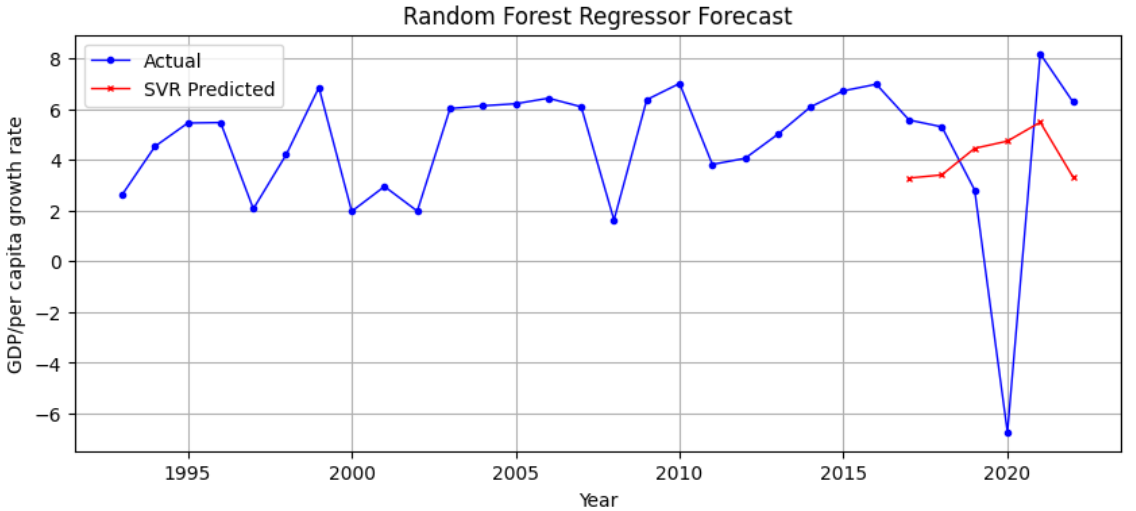


**Figure 5:** Actual data and predicted data using LSTM

# VI RANDOM FOREST REGRESSOR

# 

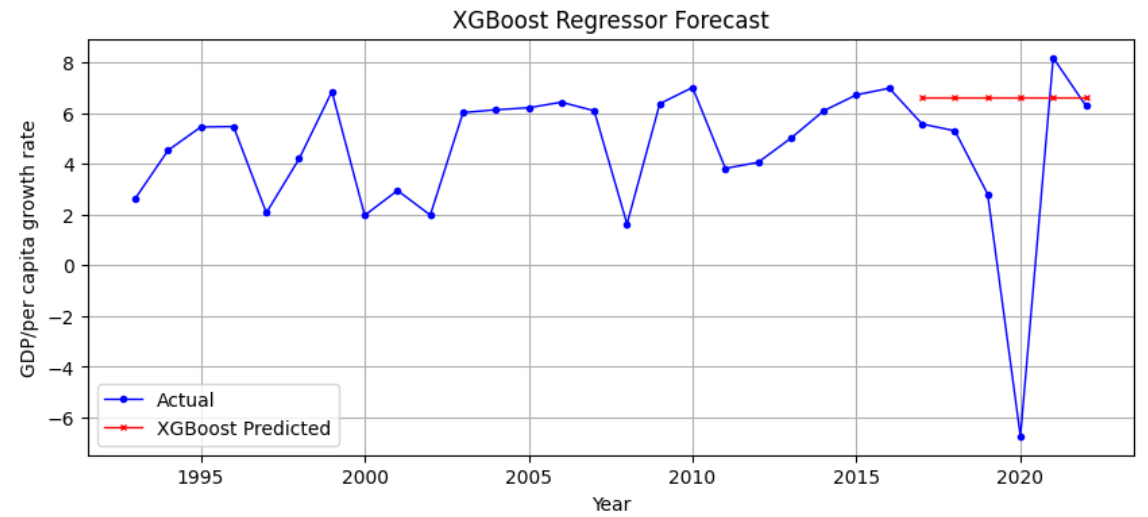
By applying lag information (previous values) as input features, the random forest regression model may be improved for time series prediction. Based on previous information, every decision tree in the ensemble has the ability to make its own forecasts, which are subsequently combined to generate the final prediction. Figure 6 shows the actual data and the support vector model-predicted data. The RSME value obtained for the model was 5.15.



**Figure 6:** Actual data and predicted data using Random Forest Regressor

# VII XGBOOST REGRESSOR

Time-related factors (such as seasonality and lagged values) are inserted into the dataset utilizing XGBoost for time series forecasting. Hyperparameter tuning is necessary for optimum performance, as the model is shown how to project future values based on prior information. Figure 7 shows the XGBoost Model's anticipated and actual values. The RMSE value for this model was 5.75.



**Figure 5:** Actual data and predicted data using XGBoost

VIII RESULT AND CONCLUSION

Through an array of machine learning models, we examined the difficulty of forecasting India's GDP per capita. The Long Short-Term Memory (LSTM) deep learning model, XGBoost regressor, Random Forest regressor, Support Vector Regressor, and ARIMA model were compared. We discovered that the Random Forest Regressor and ARIMA forecasting models performed better than other models, exhibiting the greatest accuracy for forecasting India's GDP per capita. Our evaluation was based on the Root Mean Square Error (RMSE) measure. For economists, scholars, and policymakers who care about GDP forecasting and India's economic progress, this paper makes a significant contribution to the field of predictive modeling in economics.

VIII REFERENCE

[1] Tanvi Gharte, Himani Patil, Soniya Gawade (2022). GDP Prediction and Forecasting using Machine Learning, International Research Journal of Engineering and Technology (IRJET), 2002

[2] Bhanu Pratap and Shovon Sengupta (2019). Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts? Working Paper Series, Reserve Bank of India, 2019

[3] Dwarakanath G V, Shivakumara T, Shraddha (2022). World’s GDP Prediction Using Machine Learning. Journal of Positive School Psychology - (2022, Vol. 6, No. 4, 1039-1047).

[4] Sven Heeren (2021). Forecasting Inflation using Machine Learning for an Emerging Economy. Erasmus University of Rotterdam- Bachelor Thesis Econometrics and Operations Research, 2021

[5] Kriti Mahajan & Anand Srinivasan(2018). Inflation Forecasting in Emerging Markets: A

Machine Learning Approach. Issues in Central Banking, Chapter 2, BIS Publications

[6] Gour Sundar Mitra Thakur, Rupak Bhattacharyya, Seema Sarkar Mondal. Artificial Neural

Network Based Model for Forecasting of Inflation in India. Fuzzy Information and Engineering, March 2016

[7] Gustavo Silva Araujo, Wagner Piazza Gaglianone (2020). Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models

[8] Sune karlsson, Farrukh Javed, JUNE, 2016. Modeling and Forecasting Unemployment Rate In Sweden using various Econometric Measures. Meron Desaling (86/01/30)